The Trust region affine interior point algorithm for convex and nonconvex quadratic programming

J. Frederic Bonnans, Mustapha Bouhtou

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

The Trust Region Affine Interior Point Algorithm for Convex and Nonconvex Quadratic Programming

Joseph Frédéric BONNANS
Mustapha BOUHTOU

N° 1947
Juin 1993

PROGRAMME 5

Traitement du signal,
automatique et
productique

Rapport
de recherche

1993

The trust region affine interior point algorithm for convex and nonconvex quadratic programming

La méthode de points intérieurs avec région de confiance pour la programmation quadratique convexe et non convexe

Joseph Frédéric Bonnans and Mustapha Bouhtou
INRIA, B.P. 105, 78153 Rocquencourt, France

July 16, 1993

Abstract

We study from a theoretical and numerical point of view an interior point algorithm for quadratic QP using a trust region idea, formulated by Ye and Tse. We show that, under some nondegeneracy hypothesis, the sequence of points converges to a stationary point at a linear rate. We obtain also an asymptotic linear convergence rate for the cost that depends only on the dimension of the problem. Then we show that, provided some modifications are added to the basic algorithm, the method has a good numerical behaviour.

Résumé

Nous étudions du point de vue théorique et numérique un algorithme de points intérieurs pour la programmation quadratique convexe et non convexe. Dans cet algorithme formulé par Ye et Tse, on utilise l'idée de région de confiance. Nous montrons sous une hypothèse de non dégénérescence, que l'algorithme converge linéairement vers un point stationnaire. Nous obtenons aussi un taux de convergence asymptotiquement linéaire du critère. Celui-ci ne dépend que de la dimension du problème. Avec quelques modifications de l'algorithme original, nous montrons que la méthode a un bon comportement numérique.

Key Words Interior point methods, quadratic programming, trust region, linesearch.

1 Introduction

Since Karmarkar published his polynomial projective algorithm for linear programming, the version using a simple affine transformation, commonly known as the affine scaling method, immediatly attracted interest from several researchers. Its mechanism consists of minimizing the cost over a sequence of ellipsoids whose shapes depend advantageously on the distances from the current interior feasible point to the faces of the feasible polyhedron. This method was initially developped for linear problems. It had already been proposed in 1967 by Dikin [10] who proved its global convergence under a primal nondegeneracy assumption and with scaled unit displacement stepsize. Independently Barnes [3] and Vanderbei et al [29] rediscovered Dikin's method and proved its convergence for large step size assuming primal and dual nondegeneracy hypothesis. Tseng and Luo [25] proved that the method converges if a step length is close to 2^{-L} where L is the bit size in the input.

The best results in linear programming were recently found by Tsuchiya [25] who proved the global convergence of the algorithm with one eighth scaled stepsize, for degenerate problems, and by Tsuchiya and Muramatsu [27] who proved the same result but with two third step length. They have used a local Karmarkar potential function to prove their strong results. More recently, Tsuchiya and Monteiro [28] proved that a variant of the long-step affine scaling algorithm can have a two-step superlinear convergence property for general linear programming.

The boundary behaviour of the method was studied by Megiddo and Shub [19]. They showed that the path of the continuous version of the method visits the neighborhood of all the vertices of the Klee-Minty cube (if the starting point is chosen close to the boundary) and then the complexity may not be polynomial. However, the polynomial time complexity of the discrete algorithm is still an open question. Monteiro, Adler and Resende [20] showed polynomial time complexity of the primal-dual version when the starting point is close to the central path and using very short steps. In spite of this, several experimental results [1] show the good practical behavior of the algorithm. Later, Ye and Tse [30] extended the algorithm to convex and quadratic programs. As in the linear case, all theoretical work for the convex quadratic problem, deals essentially with the global convergence question. Assuming the primal nondegeneracy condition, Ye and Tse [30] proved global convergence. With the same technique as for linear programming, Tsuchiya [27] also proved global convergence under only the dual nondegeneracy hypothesis. The best known result was proved by Sun [24] with the step-size close to 2^{-L} (L is the bit size in the input). He proved global convergence without nondegeneracy assumptions, but his displacement step size makes the algorithm impractical. Recently, Ye [31] also published some results on the affine scaling algorithm for nonconvex quadratic programming and studied in particular the complexity of the minimization of a quadratic function over an ellipsoid. Gonzaga and Carlos [12] proved global convergence of the first order version of the affine scaling algorithm for linearly constrained convex problems under a primal nondegeneracy assumption. In brief, for the trust region affine scaling algorithm in quadratic programming case, the following three hypothesis are fundamental to the most of the previous work:

1. The objective function has to be strictly convex

2. The scaled step size must be strictly smaller than one

3. A nondegeneracy condition holds.

The main aim of this paper is to analyze the convergence of the affine scaling method applied to more general quadratic programs without both the strict convexity condition and the restrictive assumption 2. For this, we have used advantageously some properties of the trust region idea.

The paper is organized as follows. First of all, we prove, under hypotheses 2 and 3 and using an hypothesis weaker than 1, that the sequence of points generated by the affine scaling algorithm with trust region converges to a stationary point at a linear rate. We show that this point is the global optimum over the polyhedron's face which contains it as an interior point. In the convex case, we prove the global optimality for any limit point of the algorithm. Convergence of the sequence of dual estimates is also established. We also obtain an asymptotic linear convergence rate for the cost that depends only on the dimension of the problem. Next, relaxing the choice of steps size, and adding an extra linesearch to basic algorithm, we prove the same convergence results. Finally, we present our numerical results which show the good behavior of the algorithm in practice.

2 The basic algorithm and its theoretical properties

Given a quadratic cost

$$f(x) := c^t x + \frac{1}{2} x^t Q x$$

with $c \in \mathbb{R}^n$ and Q a $n \times n$ symmetric matrix, we consider the problem

$$(P) \qquad \min f(x) \; ; \; Ax = b \; ; \; x \geq 0,$$

with A a $p \times n$ matrix. As we do not suppose Q to be positive, problem (P) is in general nonconvex.We denote by $A^{-1}b$ the set $\{x \in \mathbb{R}^n \; ; \; Ax = b\}$ and by F the set of feasible points, that is

$$F := \{x \in \mathbb{R}^n : Ax = b, \; x \geq 0\},$$

and by $\overset{o}{F}$ the set of "strictly feasible" points, i.e.

$$\overset{o}{F} := \{x \in \mathbb{R}^n \; ; \; Ax = b, \; x > 0\}.$$

In the sequel we assume that F is bounded and $\overset{o}{F}$ is non empty. Define $X_k := diag(x^k)$. We consider the following algorithm, proposed by Ye and Tse [30] :

Algorithm 1

0) Choose $x^0 \in \overset{o}{F}$, $\delta \in (0,1)$; $k \leftarrow 0$.

1) Compute x^{k+1} solution of

$$(SP) \qquad \min f(x) \; ; \; Ax = b \; ; \; (x - x^k)^t X_k^{-2}(x - x^k) \leq \delta^2.$$

if $(x^{k+1} - x^k)^t X_k^{-2}(x^{k+1} - x^k) < \delta^2$, stop
2) $k \leftarrow k + 1$. Go to 1). □

The non-trivial part of the algorithm consists in solving (SP) at each step. The region

$$E_k := \{x \in \mathbb{R}^n \; ; \; (x - x^k)^t X_k^{-2}(x - x^k) \leq \delta^2\}$$

can be interpreted as the Euclidian ball with radius δ after scaling, i.e, making the change of variable $x \mapsto X_k^{-1}x$ that maps x^k to $e := (1 \cdots 1)^t$. As $\delta < 1$ it follows that $E_k \cap \{x \in \mathbb{R}^n \; ; \; Ax = b\}$ is included in $\overset{o}{F}$, hence the algorithm generates a sequence of strictly feasible points. Note that (SP) has a quadratic constraint, hence it cannot in general be solved exactly : this will be discussed in section 3. In order to state our main results we need some definitions and hypotheses. Given $x \in F$, by $I(x)$ we define

$$I(x) := \{i \in \{1, \ldots, n\} \; ; \; x_i = 0\}.$$

To $I \subset \{1, \ldots, n\}$ we associate the optimization problem:

$$(P)_I \qquad \min f(x) \; ; \; Ax = b \; ; \; x_I = 0.$$

We state for future reference the optimality system of $(P)_I$.

$$(OS)_{I(x)} \quad \begin{cases} \nabla f(x) + A^t\lambda - \mu = 0 \\ Ax = b \\ x_I = 0 \\ \mu_i = 0 \; \forall i \notin I(x). \end{cases}$$

As $(\mathrm{P})_I$ is a quadratic problem with only linear equality constraints, its set of solutions is a (possibly empty) affine space.

Some of our results will use two hypotheses. The first one is

$(H1)$ for all $I \subset \{1, \ldots, n\}$, problem $(\mathrm{P})_I$ has at most one solution in F.

Note that (H1) is satisfied if Q is positive (or negative) definite.

We will also use a constraint qualification hypothesis for the limit-points of $\{x^k\}$: $\bar{x} \in F$ is said to be qualified if the following primal nondegeneracy hypothesis holds:

$$(H2) \qquad \begin{cases} \text{the relation } (A^t\lambda)_i = 0, \ \forall i \notin I(\bar{x}) \\ \text{implies that } \lambda = 0. \end{cases}$$

Note that (H2) is equivalent to

$$A^t\lambda = \sum_{i \in I(\bar{x})} \mu_i e_i$$

implies $\lambda = 0$; in this case $\mu_i = 0$ for all $i \in I(\bar{x})$; i.e. (H2) is no more that the hypothesis of linear independence of the gradients of active constraints. Hypothesis (H1) and (H2) will be useful for establishing the convergence to a point satisfying the first order optimality system. In the case of a convex cost the convergence of the cost to its optimal value can be established by assuming merely (H2).

We now state the main result of this section :

Theorem 2.1 *Let $\{x^k\}$ be a sequence generated by Algorithm 1. Then :*

(i) *If at a given step k, it happens that $(x^{k+1}-x^k)^t X_k^{-2}(x^{k+1}-x^k) < \delta$, then x^{k+1} is a global solution of (P) and $x^\ell = x^{k+1}$ for all $\ell > k$.*

(ii) *Any limit point $\bar{x}$ of $\{x^k\}$ is a solution of $(P)_{I(\bar{x})}$.*

(iii) *If (H1) holds, the sequence $\{x^k\}$ converges to some $\bar{x}$. If in addition (H2) holds then $\bar{x}$ satisfies the first-order optimality system of (P), i.e.*

$$\begin{aligned} &\nabla f(\bar{x}) + A^t\bar{\lambda} - \bar{\mu} = 0, \\ &A\bar{x} = b, \\ &\bar{x} \geq 0, \ \bar{\mu} \geq 0, \ \bar{x}^t\bar{\mu} = 0. \end{aligned} \tag{2.1}$$

(iv) *If f is convex and (H2) holds, then any accumulation point of the sequence (x^k) is an optimal solution of problem (P).*

The proof uses the optimality system of (SP), that is stated below and is a simple extension of the known result for problems without equality constraints, see [7].

Lemma 2.1 *The point x^{k+1} solution of (SP) is characterized by the existence of $\lambda^{k+1} \in \mathbb{R}^p$, $\nu_k \geq 0$ such that*

$$\nabla f(x^{k+1}) + A^t\lambda^{k+1} + \nu_k X_k^{-2}(x^{k+1} - x^k) = 0, \tag{2.2}$$

$$Ax^{k+1} = b \tag{2.3}$$

$$\left.\begin{aligned} &\nu^k \geq 0, \ (x^{k+1} - x^k)^t X_k^{-2}(x^{k+1} - x^k) \leq \delta^2 \\ &\nu_k[(x^{k+1} - x^k)^t X_k^{-2}(x^{k+1} - x^k) - \delta^2] = 0 \end{aligned}\right\} \tag{2.4}$$

$$d^t(Q + \nu_k X_k^{-2})d \geq 0, \ \forall d \in N(A) := \{x \in \mathbb{R}^n \ ; \ Ax = 0\}. \tag{2.5}$$

Remark 2.1 *That (2.2),(2.3) and (2.5) hold is equivalent to : the function*

$$\varphi_k(x) := f(x) + \frac{\nu_k}{2}(x - x^k)^t X_k^{-2}(x - x^k)$$

is convex on $A^{-1}b$ *and attains its minimum on* $A^{-1}b$ *at* x^{k+1}.

The essential ingredient of the proof of Theorem 2.1 is

Proposition 2.1 *Let* $\{x^k\}$ *be a sequence generated by Algorithm 1, and* (ν_k, λ^k) *the associated multipliers. Then*

(i) $\delta^2 \sum_{\ell=0}^{k} \nu_\ell \leq 2(f(x^0) - f(x^{k+1}))$,

(ii) $\delta \sum_{\ell=0}^{k} \|X_\ell(\nabla f(x^{\ell+1}) + A^t\lambda^{\ell+1})\| \leq 2(f(x^0) - f(x^{k+1}))$,

(iii) $(x^k - x^{k+1})^t Q(x^k - x^{k+1}) \to 0$,

(iv) *If* $(\bar{x}, \hat{x})$ *is a limit point of* (x^k, x^{k+1}) *then* $I(\bar{x}) = I(\hat{x})$.

Proof

(i) Using Remark 2.1, it follows

$$\varphi_k(x^{k+1}) \leq \varphi_k(x^k) = f(x^k) \tag{2.6}$$

and

$$f(x^{k+1}) + \nu_k \frac{\delta^2}{2} = \varphi_k(x^{k+1}), \tag{2.7}$$

hence

$$\delta^2 \nu_k \leq 2(f(x^k) - f(x^{k+1}));$$

point (i) follows.

(ii) From (2.2) we deduce

$$X_k[\nabla f(x^{k+1}) + A^t\lambda^{k+1}] = -\nu_k X_k^{-1}(x^{k+1} - x^k).$$

From $\|X_k^{-1}(x^{k+1} - x^k)\| = \delta$ and (i) we deduce (ii).

(iii) As φ_k is quadratic it follows that

$$\varphi_k(x^k) = \varphi_k(x^{k+1}) + \nabla\varphi_k(x^{k+1})^t(x^k - x^{k+1}) + \frac{1}{2}(x^k - x^{k+1})^t\nabla^2\varphi_k(x^{k+1})(x^k - x^{k+1}).$$

By (2.2):

$$\nabla\varphi_k(x^{k+1})^t(x^k - x^{k+1}) = -(\lambda^{k+1})^t A(x^k - x^{k+1}) = 0.$$

From (2.4) :

$$\nu_k(x^k - x^{k+1})^t X_k^{-2}(x^k - x^{k+1}) = \nu_k\delta^2,$$

hence

$$f(x^k) = \varphi_k(x^k) = \varphi_k(x^{k+1}) + \frac{1}{2}(x^k - x^{k+1})^t Q(x^k - x^{k+1}) + \frac{\nu_k}{2}\delta^2.$$

This and (2.7) imply

$$f(x^k) - f(x^{k+1}) - \nu_k\delta^2 = \frac{1}{2}(x^k - x^{k+1})^t Q(x^k - x^{k+1}).$$

By (i), the monotonic decrease of f and the fact that $\inf(P) > -\infty$, the left hand side goes to 0 when $k \to \infty$, and the result follows.

(iv) As $x^{k+1} \in E_k$, it follows that $(1-\delta)x_i^k \leq x_i^{k+1} \leq (1+\delta)x_i^k$, $i = 1, \ldots, n$, henceforth for a converging subsequence, $x_i^{k+1} \to 0$ iff $x_i^k \to 0$; the result follows. □

Remark 2.2 *When Q is positive definite, Ye [30] and Tsuchiya [27] used property iii) of Prop. 2.1 to prove the convergence of the sequence (x^k). Here, we shall rather use the more general hypothesis (H1).*

As in the analysis of C. Gonzaga and L. Carlos [12], we shall use the following interesting result based on a well known property of convex analysis, for which we refer to O.L. Mangasarian [18] and J. Burke and Ferris [8].

Lemma 2.2 *If f is a convex function then the gradient function $\nabla f(.)$ is constant on the optimal solution set of $(P)_{I(\bar{x})}$.*

Proof of theorem 2.1

(i) If it happens that $(x^{k+1} - x^k)^t X_k(x^{k+1} - x^k) < \delta$, using Lemma 2.1 and Remark 2.1 we deduce that $\nu_k = 0$ and that $\varphi_k(x) = f(x)$ attains its minimum on $A^{-1}b$ at x^{k+1}. It follows that x^{k+1} is a solution of (P).

(ii) Let $\bar{x}$ be a limit-point of $\{x^{k+1}\}$. Denote $I := I(\bar{x})$, $\bar{I} := \{1, \ldots, n\} - I$. From Prop 2.1 (ii) we get for the given subsequence

$$(\nabla f(x^{k+1}) + A^t \lambda^{k+1})_{\bar{I}} \to 0. \tag{2.8}$$

Define $G := \{(A^t\lambda)_{\bar{I}},\ \lambda \in \mathbb{R}^p\}$. Then G is a linear subspace ; from (2.8) it follows that $dist((\nabla f(x^{k+1}))_I, G) \to 0$; as G is closed there exist some $\bar{\lambda}$ such that $(\nabla\ f(\bar{x})\ +\ A^t\ \bar{\lambda})_{\bar{I}}\ =\ 0$. From this it follows that $\bar{x}$ satisfies the first order optimality system of $(\mathrm{P})_{I(\bar{x})}$.

Now let $d \in \ker A$ be such that $d_I = 0$ (the set of such d's is possibly $\{0\}$). Then passing to the limit in (2.5), and reminding that ν_k converges to zero by Prop 2.1 (i), it follows that $\nu^k d^t X_k^{-2} d \to 0$, hence $d^tQd \geq 0$; now if x is feasible for $(\mathrm{P})_I$, then $d := x - \bar{x}$ is in $\ker A$ and $d_I = 0$ hence

$$\begin{aligned} f(x) &= f(\bar{x}) + \nabla f(\bar{x})^t d + \frac{1}{2} d^t Q d \\ &= f(\bar{x}) + \frac{1}{2} d^t Q d \geq f(\bar{x}) \end{aligned}$$

which proves (ii).

(iii) From Prop 2.1 (iv), point (ii) and (H1) we deduce that if $(\bar{x}, \hat{x})$ is limit-point of (x^k, x^{k+1}) then $\bar{x} = \hat{x}$. In particular $\|x^{k+1} - x^k\| \to 0$ which implies that the set of limit points of $\{x^k\}$ is connected. Using (ii) and (H1) again we deduce that the whole sequence $\{x^k\}$ converges to some $\bar{x}$.

We now prove that if (H2) also holds then $\bar{x}$ satisfies the optimality system of (P). From Prop 2.1 (ii) we deduce

$$(\nabla f(x^{k+1}) + A^t\lambda^{k+1})_i \to 0, \quad i \notin I(\bar{x}),$$

hence with (H2) $\lambda^k \to \bar{\lambda}$ such that $(\nabla f(\bar{x}) + A^t\bar{\lambda})_i = 0$ for all $i \notin I(\bar{x})$, hence optimality system of $(OS)_{I(\bar{x})}$ is satisfied:

$$\begin{aligned} &\nabla f(\bar{x}) + A^t\bar{\lambda} - \bar{\mu} = 0, \\ &A\bar{x} = b, \\ &\bar{\mu}_i = 0,\ \forall i \notin I(\bar{x}). \end{aligned}$$

As $A\bar{x} = b$ it remains to prove that $\bar{\mu}_I \geq 0$. From (2.2) and the convergence of $\{\lambda^k\}$ we deduce that $\bar{\mu} = \lim \nu_k (X_k^{-2}(x^k - x^{k+1})$. If $i \in I(\bar{x})$ then $x_i^k \to 0$, hence $x_i^{k+1} \leq x_i^k$ at least for a subsequence, and it follows that $\bar{\mu}_i \geq 0$.

(iv) Denote by μ^k the dual estimate term given by (2.2): $\mu^k := \nabla f(x^{k+1}) + A^T\lambda^{k+1}$. We shall first prove that the sequence (μ^k) converges : it is well known that hypothesis (H2) implies that for any limit point x of (x^k), the matrix (AX^2A^T) is not singular, where $X = \text{diag}(x)$. As $\mu^k = [I - A^T(AX_k^2A^T)^{-1}AX_k^2]\nabla f(x^{k+1})$, we deduce that the sequence (μ^k) is bounded.

Now let $\bar{x}$ be a limit-point of a sub-sequence (x^{k+1}). If x^* is another limit-point such that $I(\bar{x}) \subset I(x^*)$ then using (ii) of the Theorem 2.1, the fact that $f(\bar{x}) = f(x^*)$ and Lemma 2.2, we deduce that $\nabla f(\bar{x}) = \nabla f(x^*)$. Again the nondegeneracy hypothesis (H2) implies that for any subsequence $(x^{k'})$ such that $x^{k'+1} \to x^*$, we have that $\lim_k \mu^k = \lim_{k'} \mu^{k'}$. This and the fact that that the set of faces of F is finite imply that the set of the accumulation points of the dual estimates sequence is also finite.

Now from (iv) of proposition 2.1 and proceeding as above with $k' = k+1$, it is easy to see that $\lim \nabla f(x^{k'+1}) = \nabla f(\bar{x})$ and hence $\mu^{k+1} - \mu^k$ converges to zero. Hence the set of accumulation points of (μ^k) is connected and finite. This implies that the sequence (μ^k) converges to some $\bar{\mu}$.

Now we are ready to prove (iv).

Let $\bar{x}$ an accumulation point of the sequence x^k, and $\bar{\lambda}$ the limit of λ^{k+1}. Then $(OS)_{I(\bar{x})}$ is satisfied with multipliers $\bar{\lambda}$ and $\bar{\mu}$.

As in the proof of (iii) it remains to prove that $\bar{\mu}_I \geq 0$. Assume by contradiction that there exists $i \in I$ such that $\bar{\mu}_i < 0$ then by complementarity $\bar{x}_i = 0$ and the convergence of (μ^k) implies that $\nu^k \dfrac{x_i^k - x_i^{k+1}}{(x_i^k)^2} < 0$ for all k large enough. Hence $\bar{x}_i = 0$ and $x_i^k < x_i^{k+1}$ for all k large enough. This is a contradiction and so $\bar{\mu}_I \geq 0$. □

Remark 2.3 i) *In the statement of (iii) of Theorem 1, instead of (H1) we can obviously assume only that for any limit point $\bar{x}$ of the sequence $\{x^k\}$, $P_{I(\bar{x})}$ has at most one solution in F.*

ii) *As it is showen in the proof of (iv) of Theorem 1, the sequence of dual estimates $\mu^k = \nabla f(x^{k+1}) + A^T\lambda^{k+1}$ converges when the objective function is convex. This important theoretical property may be well exploited in practice since it allows during the iterations of the algorithm to identify the strongly active constraints at optimality. By strongly active constraint we mean an index i in $\{1, \ldots, n\}$ such that $\bar{\mu}_i > 0$. Indeed, if a component i is strongly active then $\mu_i^k > 0$ for all k large enough and from the fact that $\mu^k = \nu^k X_k^{-2}(x^k - x^{k+1})$ we deduce that the sequence (x_i^k) strictly decreases for k large enough.*

Note that C. Gonzaga and Carlos [12], under primal nondegeneracy assumption also proved the optimality conditions for the first order version of the affine scaling method applied to linearly constrained convex programs. They did not prove the convergence of the sequence of the dual estimates. However, one can easily proceed as in our proof of (iv) of Theorem 1, in order to also prove the convergence of the dual estimates generated by the first order affine scaling method.

Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a solution of the optimality system (2.1). We say that $\bar{x}$ satisfies the strict complementarity hypothesis if

$$\bar{x}_i > 0 \text{ or } \bar{\mu}_i > 0, \ \forall i \in \{1, \ldots, n\}. \tag{2.9}$$

We now analyze the rate of convergence of the algorithm. We note that even in the case of a linear cost, the known results deal only with the asymptotic rate of convergence of the cost. We generalize these results in point (i) of Thm 2.2. We also give a result concerning the speed of convergence of $\{x^k\}$. We denote

$$\|x\|_k := \sqrt{\sum (x_i/x_i^k)^2} = x^t X_k^{-2} x.$$

Theorem 2.2 *Let $\{x^k\}$ be generated by Algorithm 1. Then :*

(i) *For all $x^* \in F$ such that $f(x^*) < f(x^{k+1})$ then*

$$f(x^{k+1}) - f(x^*) \leq \left(1 - \frac{\delta^2}{\|x^k - x^*\|_k^2}\right)(f(x^k) - f(x^*)). \tag{2.10}$$

In particular, if x^k converges to $\bar{x}$ then for some $\varepsilon_k \to 0$:

$$f(x^{k+1}) - f(\bar{x}) \leq \left(1 - \frac{\delta^2}{card(I(\bar{x})) + \varepsilon_k}\right)(f(x^k) - f(\bar{x})), \tag{2.11}$$

and in particular

$$f(x^{k+1}) - f(\bar{x}) \leq \left(1 - \frac{\delta^2}{n + \varepsilon_k}\right)(f(x^k) - f(\bar{x})). \tag{2.12}$$

(ii) *In addition, if (H1), (H2) and the strict complementarity hypothesis (2.9) hold then*

$$\sum_k \|x^k - \bar{x}\| + \sum_k \|\lambda^k - \bar{\lambda}\| < +\infty. \tag{2.13}$$

Proof

(i) As φ_k attains its minimum in $A^{-1}b$ at x^{k+1} it follows that

$$\begin{aligned}
\varphi_k(x^{k+1}) &\leq \varphi_k(x^k) = f(x^k), \\
\varphi_k(x^{k+1}) &\leq \varphi_k(x^*) = f(x^*) + \frac{\nu_k}{2}\|x^* - x^k\|_k^2.
\end{aligned}$$

We obtain that for all $\theta \in [0, 1]$

$$\varphi_k(x^{k+1}) \leq (1 - \theta)\varphi_k(x^k) + \theta\varphi_k(x^*),$$

hence

$$\varphi_k(x^{k+1}) - \varphi_k(x^*) \leq (1-\theta)[f(x^k) - \varphi_k(x^*)],$$

therefore

$$f(x^{k+1}) - f(x^*) \leq (1-\theta)[f(x^k) - f(x^*)] + \frac{\nu_k}{2}[\theta\|x^* - x^k\|_k^2 - \delta^2].$$

We note that, as $f(x^*) < f(x^{k+1})$, x^* must be outside the ellipsoid E_k.

Choosing $\theta := \delta^2/\|x^* - x^k\|_k^2 < 1$ we deduce (2.10), relations (2.11) and (2.12) easily follow.

(ii) Define $I := I(\bar{x})$. From Prop 2.1 (ii) and (2.9) we deduce, as $x^k \to \bar{x}$, when $i \notin I$

$$\sum_{k=0}^{\infty} |(\nabla f(x^{k+1}) + A^t\lambda^{k+1})_i| < \infty. \tag{2.14}$$

Now from Prop 2.1 (ii) and (2.9) again :

$$\sum_{k=0}^{\infty} x_i^k < \infty \text{ for all } i \in I. \tag{2.15}$$

Define $\eta^{k+1} \in \mathbb{R}^n$ by

$$\eta_i^{k+1} := \begin{cases} (\nabla f(x^{k+1}) + A^t\lambda^{k+1})_i & if \quad i \notin I \\ 0 & if \quad i \in I. \end{cases}$$

It appears that (x^{k+1}, λ^{k+1}) are primal-dual solutions of the optimality system of

$$\min_x f(x) - \eta^t x \; ; \; Ax = b \; ; \; x_I = x_I^{k+1},$$

or equivalently $(x^{k+1} - \bar{x}, \lambda^{k+1} - \bar{\lambda})$ is solution of the linear system in (x, λ) :

$$\begin{cases} (Qx + A^t\lambda)_{\bar{I}} &= \eta_{\bar{I}}^{k+1} \\ Ax &= 0 \\ x_I &= x_I^{k+1}. \end{cases} \tag{2.16}$$

We claim that this system is invertible. Indeed consider a solution to

$$\begin{cases} (Qx + A^t\lambda)_{\bar{I}} &= 0 \\ Ax &= 0, \\ x_I &= 0, \end{cases}$$

hence $0 = x^t(Qx + A^t\lambda)x = x^tQx$. Now $\bar{x} + \rho x$ is feasible for $(P)_I$ and

$$f(\bar{x} + \rho x) = f(\bar{x}) + \rho\nabla f(\bar{x})^t x + \frac{\rho^2}{2}x^tQx = f(\bar{x});$$

by (H1) x must be null, hence $\lambda = 0$ by (H2) which proves the invertibility of this linear system. Now by invertibility of (2.16) and (2.14), (2.15) there exists $K > 0$ such that

$$\sum_{k=0}^{\infty}(\|x^k - \bar{x}\| + \|\lambda^k - \bar{\lambda}\|) \leq K \sum_{k=0}^{\infty}(\|\eta^k\| + \|x_J^k\|) < \infty.$$

□

Remark 2.4 *i) Recently Ye [31] also proved a similar result on the convergence rate for the cost. However our proof is quite different and simpler than the one of Ye.*

ii) In the linear programming case, the two-third sep affine scaling algorithm converges to an optimal point which satisfies the strict-complementarity property. Unfortunately this may not be the case in quadratic programming as it may happen that no solution satisfies (2.9) as in the following trivial example

$$\begin{cases} \min & x_1^2 \\ x_1 \geq 0, & x_2 \geq 0 \end{cases}$$

3 The extended algorithm

So far we have analyzed Algorithm 1, supposing that the solution of (SP) could be computed exactly : this is not the case, however, as problem (SP) is nonlinear. However, guessing a value for the multiplier ν associated with the nonlinear constraint it is possible to solve exactly, whenever it has a solution, the problem of minimizing the associated Lagrangian :

$$(Q)_\nu \qquad \min \psi_\nu^k(x) := f(x) + \frac{\nu}{2}(x - x^k)^t X_k^{-2}(x - x^k) \quad s.t. \quad Ax = b.$$

We denote a solution of $(Q)_\nu$ (whenever it exists) by x_ν^k.

As X_k^{-2} is positive definite, there exists a treshold value $\bar{\nu}_k \geq 0$ (we do not consider negative values of ν) such that $\psi_\nu^k(x)$ is convex on $N(A)$, if and only if $\nu \geq \bar{\nu}_k$. Also by (2.5) this $\bar{\nu}_k$ satisfies $\bar{\nu}_k \leq \nu_k$. By classical argument [7], [21] one can prove that the function $\nu \rightarrow (x_\nu^k - x^k)^t X_k^{-2}(x_\nu^k - x^k)$ is strictly decreasing on $]\bar{\nu}_k, \infty[$ when x^k is not a stationary point. Hence ν_k can be computed efficiently within a given precision at least by a simple dichotomic procedure. We deal with this at paragraph 4.

In order to take in to account the fact that (SP) cannot be solved exactly, but that the solution of the trust region subproblem can be computed for a number of values of the trust region close to δ, we allow the possibility for δ to vary at each iteration. Also we add the possibility of a linesearch in the direction computed by the subproblem. This linesearch does not give any new theoretical property, but it proved very effective in our numerical tests. The algorithm is as follows :

Algorithm 2

0) Choose $x^0 \in \overset{\circ}{F}$; $\delta \in (0,1)$, $k \leftarrow 0$.

1) Compute $\hat{x}^{k+1}$ solution for some $\delta_k > \delta$ of

$$(SP2) \qquad \min f(x) \; ; \; Ax = b \; ; \; (x - x^k)^t X_k^{-2}(x - x^k) \leq \delta_k^2,$$

the parameter δ_k being such that $\hat{x}^{k+1} > 0$ (hence it is possible that $\delta_k > 1$).
2) Linesearch : denote $d_k := \hat{x}^{k+1} - x^k$.
Fix $\gamma_k \geq 1$ such that $x^k + \gamma_k d^k > 0$.
Compute $\rho_k = \arg\min\{f(x^k + \rho d^k),\ \rho \in [1, \gamma_k]\}$
3) $k \leftarrow k + 1$, go to 1). □

In the analysis we will see that it is useful to have some bounds on γ_k. Let us remark that in interior point algorithms a common rule when computing a sequence of vectors $\{x^k\}$ with positive components is to impose that the following relation holds : $x_i^{k+1} \geq \theta x_i^k$ for some $\theta \in (0,1)$; i.e : $1 + \rho_k d_i^k / x_i^k \geq \theta$. However we will rather use here the converse relation $x_i^{k+1} \leq \theta^{-1} x_i^k$. Hence we will choose γ_k in Algorithm 2 such that

$$0 < 1 + \rho_k d_i^k / x_i^k \leq \theta^{-1}. \tag{3.17}$$

$\delta_k > \delta$ in Algorithm 2 means that the size of the region in which $\hat{x}^{k+1}$ is computed is not too small.

Remark 3.1 *Indeed (3.17) gives a bound on γ_k. Excluding the trivial case $(x^k - \hat{x}^{k+1}) X_k^{-2}(x^k - \hat{x}^{k+1}) < \delta_k^2$, we deduce from $\sum_i (d_i^k / x_i^k)^2 = \delta_k^2 > \delta^2$ that $|d_i^k| / x_i^k > \delta / \sqrt{n}$ for at least some i. Now by (3.17)*

$$-1 \leq \rho_k d_i^k / x_i^k \leq \theta^{-1} - 1$$

hence

$$\rho_k |d_i^k| / x_i^k \leq \max(\theta^{-1} - 1, 1) \leq \theta^{-1}$$

It follows that $\rho_k \leq \dfrac{1}{\delta\theta}\sqrt{n}$.

That is, if (3.1) holds we may assume that $\gamma_k \leq \dfrac{1}{\delta\theta}\sqrt{n}$.

Remark 3.2

i) *(SP2) always has a solution.*

ii) *We can check that δ_k is always bounded from above. Indeed, let $\bar{x}$ be a limit-point of $\{x^k\}$. We find that*

$$\delta_k^2 \leq \|\bar{x} - x^k\|_k^2 \rightarrow card\,(I(\bar{x}))$$

otherwise $\bar{x}$ will be a solution of (SP2) and consequently problem (P) is convex and $\bar{x}$ its optimal solution. Hence, $\limsup \delta_k \leq \sqrt{n}$.

Theorem 3.1 *Let $\{x^k\}$ be a sequence generated by Algorithm 2. We assume that (3.17) holds. Then :*

(i) *If at a given step k, it happens that $(\hat{x}^{k+1} - x^k)^t X_k^{-2}(\hat{x}^{k+1} - x^k) < \delta_k$, then $\hat{x}^{k+1}$ is a global solution of (P).*

(ii) *Any limit point $\bar{x}$ of $\{x^k\}$ is a solution of $(P)_{I(\bar{x})}$.*

(iii) *If (H1) holds, the sequence x^k converges to some $\bar{x}$. If in addition (H2) holds then $\bar{x}$ satisfies the first-order optimality system of (P), i.e.*

$$\begin{aligned} &\nabla f(\bar{x}) + A^t\bar{\lambda} - \bar{\mu} = 0 \\ &A\bar{x} = b \\ &\bar{x} \geq 0,\ \bar{\mu} \geq 0,\ \bar{x}^t\bar{\mu} = 0 \end{aligned} \tag{3.18}$$

(iv) *If f is convex and (H2) holds then any accumulation point of the sequence (x^k) is an optimal solution of problem (P).*

For the proof of the Theorem 3.1 we need a statement corresponding to Proposition 2.1.

Proposition 3.1 *Let x^k be a sequence generated by Algorithm 2, and (ν_k, λ^k) the associated multipliers. We assume that (3.17) holds. Then*

i) $\sum_{\ell=0}^{k}(\delta_\ell)^2\nu_\ell \leq 2(f(x^0) - f(x^{k+1}))$,

ii) $\sum_{\ell=0}^{k}\delta_\ell\|X_\ell(\nabla f(\hat{x}^{k+1}) + A^t\lambda^{\ell+1})\| \leq 2(f(x^0) - f(x^{k+1})$,

iii) $(x^k - x^{k+1})^tQ(x^k - x^{k+1}) \to 0$,

iv) *If $(\bar{x}, \hat{x})$ is limit-point of (x^k, x^{k+1}), then $I(\bar{x}) \subset I(\hat{x})$.*

Proof

i) Proceeding as in the proof of Prop 2.1 we find that

$$\delta_k^2\nu_k \leq 2(f(x^k) - f(\hat{x}^{k+1}))$$

As $f(\hat{x}^{k+1}) \geq f(x^{k+1})$ we deduce that $\delta_k^2\nu_k \leq 2(f(x^k) - f(x^{k+1}))$; point (i) follows.

ii) This can be proved as for Prop 2.1.

iii) Proceeding as in the proof of Prop 2.1 we find that $(x^k - \hat{x}^{k+1})^tQ(x^k - \hat{x}^{k+1}) \to 0$. Now by Remark 3.1, ρ_k is bounded, and

$$|(x^k - x^{k+1})^tQ(x^k - x^{k+1})| = (\rho_k)^2|(x^k - \hat{x}^{k+1})^tQ(x^k - \hat{x}^{k+1})| \to 0.$$

iv) As $\hat{x}^{k+1} \in E_k$ we have

$$x_i^{k+1} \leq (1 + \rho_k\delta_k)x_i^k.$$

Using Remarks 3.1 and 3.2 we get $x_i^{k+1} \leq \left(1 + \frac{n+1}{\delta\theta}\right)x_i^k$ for k large enough from which the conclusion follows. □

Proof of Thm 3.1

i) The same argument as in the proof of Thm 2.1 gives the result.

ii) Let $\bar{x}$ be a limit-point of x^k. If $\bar{x}$ is not solution of $P_{I(\bar{x})}$, let x^* be feasible for $P_{I(\bar{x})}$ and $f(x^*) < f(\bar{x})$. Arguing as in Thm 2.2 we get

$$f(\hat{x}^{k+1}) - f(x^*) \leq \left(1 - \frac{\delta_k^2}{\|x^k - x^*\|_k^2}\right)(f(x^k) - f(x^*)).$$

As $\delta_k \geq \delta$ and $f(x^{k+1}) \leq f(\hat{x}^{k+1})$ we get

$$f(x^{k+1}) - f(x^*) \leq \left(1 - \frac{\delta^2}{\|x^k - x^*\|_k^2}\right)(f(x^k) - f(x^*)).$$

For the considered subsequence, as $I(x^*) \supset I(\bar{x})$, we get

$$\|x^k - x^*\|_k^2 \rightarrow card\ I(x^*) + \sum_{i \notin I(x^*)} (1 - x_i^*/\bar{x}_i)^2 < \infty$$

hence $f(x^k) \rightarrow f(x^*)$, in contradiction with our hypothesis.

iii) We have by point (ii) and (H1) that if $(\bar{x}, \hat{x})$ is limit point of (x^k, x^{k+1}) then $\bar{x}$ is unique solution of $P_{I(\bar{x})}$ and $\hat{x}$ is unique solution of $P_{I(\hat{x})}$. As $I(\hat{x}) \supset I(\bar{x})$ it follows (by definition of P_I) that $\bar{x}$ is feasible for $P_{I(\hat{x})}$, and $f(\hat{x}) = f(\bar{x}) = \lim f(x^k)$. This implies $\hat{x} = \bar{x}$, and in particular $\|x^{k+1} - x^k\| \rightarrow 0$, i.e. the set of limit points of $\{x^k\}$ is connected. By (ii) and (H1) each of these limit-points is isolated. We deduce that $\{x^k\}$ converges towards some $\bar{x}$. Now $\|\hat{x}^{k+1} - x^k\| \leq \|x^{k+1} - x^k\|$ hence $\hat{x}^k$ has also limit $\bar{x}$. We obtain as in the proof of Thm 2.1

$$\begin{cases} \nabla f(\bar{x}) + A^t\bar{\lambda} - \bar{\mu} = 0, \\ A\bar{x} = b, \\ \bar{\mu}_i = 0,\ \forall i \notin I(\bar{x}) \end{cases}$$

and, thanks to (H2), $\bar{\mu} = \lim \nu_k X_k^{-2}(x^k - \hat{x}^{k+1})$. For $i \in I(\bar{x})$, $x^k \rightarrow 0$ and $x^{k+1} - x^k = \rho_k(\hat{x}^{k+1} - x^k)$ with $\rho_k > 0$ hence $x_i^k \geq \hat{x}_i^{k+1}$ for at least a subsequence, hence $\bar{\mu} \geq 0$. and then (3.18) holds.

iv) Denote by $\mu^k := (\nabla f(\hat{x}^{k+1}) + A^T\lambda^{k+1})$ the dual estimate. Using again the convexity of f and the nondegeneracy assumption, arguing as in the proof of iv) of theorem 2.1, we prove that (μ^k) is bounded and the set of its limit points is finite.

Using (ii) of Theorem 3.1 and from iv) of proposition 3.1, again as in the proof of iv) of Theorem 2.1, we deduce that $\|\mu^{k+1} - \mu^k\|$ converges to zero and hence the sequence (μ^k) converges. Thus, it is easy to show the optimality conditions which imply that iv) holds.

Remark 3.3 *Theorem 2.2 has an immediate extension to Algorithm 2.*

4 A practical algorithm and numerical results

There is no doubt that, solving problem (SP2) is the most important and hardest stage of algorithm 2. The linesearch is of course easy since the function is quadratic.

(SP2) is a classical problem which has been greatly studied by researchers. Hence one can efficiently solve it by one of the classical algorithms used to compute the displacement step in trust region methods (see Reinsch [22], Hebden [15], D.C. Sorensen [23], J.J. Moré [21], M. Gay [11]). These algorithms generally use Newton's method to compute the multiplier ν_k which in our case verifies the relation $\|x_{\nu_k} - x^k\|_k = \delta_k$, where x_{ν_k} is such that:

$$\begin{pmatrix} Q + \nu_k X^{-2} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x_{\nu_k} - x^k \\ \lambda^k \end{pmatrix} = - \begin{pmatrix} \nabla f(x^k) \\ 0 \end{pmatrix} \tag{4.19}$$

In our numerical tests, we have used instead an algorithm based on a simple dichotomic procedure. Indeed, on the one hand, we know (see the proof of Prop 2.1) that $\nu_k \in]\bar{\nu}_k, 2\delta^{-2}(f(x^k) - f(x^{k+1}))]$ where $\bar{\nu}$ and δ are as in paragraph 3.

On the other hand, the function $\nu \mapsto \|x_\nu - x^k\|_k$ is strictly decreasing on $]\bar{\nu}_k, +\infty[$. Hence, knowing the estimations of $f(x^{k+1})$ and $\bar{\nu}_k$ will be sufficient to be able to compute an estimation of the multiplier ν_k, the solution of $\|x_\nu - x^k\|_k = \delta_k$, by the iterative dichotomic on $]\bar{\nu}_k, 2\delta^{-2}(f(x^k) - f(x^{k+1}))]$. Unfortunately, to estimate the multiplier $\bar{\nu}_k$ is a hard problem especially when Q is indefinite. But we did proceed like this :

1. $\nu_{\text{inf}} := 0$, $\nu_{\text{sup}} := \dfrac{2}{\delta^2}[f(x^k) - \gamma_k]$ where γ_k is an under estimate of $f(x^{k+1})$

2. $\nu := \dfrac{1}{2}(v_{\text{inf}} + \nu_{\text{sup}})$

3. • $H_\nu := Z^T Q Z + \nu Z^T X_k^{-2} Z$, where Z is a basis of the nul space of the matrix A.
   • Solve the reduced system $H_\nu w = -Z^T \nabla f(x^k)$ by an iterative method which controls the positivity of H_ν (e.g. the Lanczos method).

   - if we detect that H_ν is indefinite then
     - stop the iterative method
     - $\nu_{\text{inf}} := \nu$ and go to 2.
   - Else $x_\nu \leftarrow x^k + Zw$
     - if $\|x_\nu - x^k\|_k \geq \delta$
       * if $x_\nu > 0$ stop
       * else $\nu_{\text{inf}} \leftarrow \nu$ and goto 2.
     - if $\|x_\nu - x^k\|_k < \delta$ then
       * if $\nu_{\text{sup}} - \nu_{\text{inf}} < \varepsilon$ where ε is a given precision then stop
       * else $\nu_{\text{sup}} := \nu$ and goto 2.

In the convex case, zero is a trivial estimation for $\bar{\nu}_k$ and stage 3 of the above procedure becames simple to implement. We have tested the performance of algorithm 2 to solve many convex quadratic problems randomly generated. All our tests were done on a SUN 4/65 computer and the algorithm was written in the language BASILE [4] developped at INRIA. All tested problems were of the form :

$$(P') \quad \begin{cases} \min f(x) \\ Ax = b \\ Bx \leq d \\ \ell \leq x \leq u \end{cases}$$

Introducing slack variables, $z := d - Bx$, $\bar{w} := u - x$, $\underline{w} := x - \ell$ and replacing x by $u - \bar{w}$, one can transform (P') to the standard form problem:

$$(P'') \quad \begin{cases} \min \frac{1}{2}\bar{w}^T Q\bar{w} - \nabla f(u)^T\bar{w} \\ A\bar{w} = Au - b \\ B\bar{w} - z = Bu - d \\ \bar{w} + \underline{w} = u - \ell \\ \bar{w} \geq 0, \ \underline{w} \geq 0, \ z \geq 0 \end{cases}$$

and apply algorithm (2) to solve it. However this has the great disadvantage of increasing the size of the problem, especially when the number of inequality constraints is larger than the number of variables. Indeed, if we denote by $Z_k := \mathrm{diag}(z^k)$, $\bar{W}_k := \mathrm{diag}(\bar{w}^k)$, $\underline{W}_k := \mathrm{diag}(\underline{w}^k)$, one will have to solve the following sub-problem $(SP3)$

$$(SP3) \quad \begin{cases} \min \frac{1}{2}\bar{w}^T Q\bar{w} - \nabla f(u)^T\bar{w} \\ A\bar{w} = Au - b \\ B\bar{w} - z = Bu - d \\ \bar{w} + \underline{w} = u - \ell \\ (\bar{w} - \bar{w}^k)^T \bar{W}_k^{-2}(\bar{w} - \bar{w}^k) + (\underline{w} - \underline{w}^k)^T \underline{W}_k^{-2}(\underline{w} - \underline{w}^k) \\ \quad +(z - z^k)^T Z_k^{-2}(z - z^k) \leq \delta_k^2. \end{cases}$$

For this, one need to solve as in (4.19) a linear system whose matrix is

$$\begin{bmatrix} Q + \nu_k \bar{W}_k^{-2} & 0 & 0 & A^T & B^T & I \\ 0 & \nu_k \underline{W}_k^{-2} & 0 & 0 & 0 & I \\ 0 & 0 & \nu_k Z_k^{-2} & 0 & -I & 0 \\ A & 0 & 0 & 0 & 0 & 0 \\ B & 0 & -I & 0 & 0 & 0 \\ I & I & 0 & 0 & 0 & 0 \end{bmatrix}$$

On the other hand, one can easily show that the ellipsoid centered at x^k in x space, given by sub-problem (SP3) after eliminating all slack variables, is

$$E_1 = \{x \in \mathbb{R}^n \ ; \ (x - x^k)^t \, \mathrm{diag}\left[\left\{\frac{1}{\bar{w}_i^k}\right\}^2 + \left\{\frac{1}{\underline{w}_i^k}\right\}^2\right](x - x^k) + (x - x^k)^T B^T Z_k^{-2} B(x - x^k) \leq \delta_k^2\}$$

Therefore, (SP3) is equivalent to the following sub-problem:

$$\min f(x)\ ;\ Ax = b,\ x \in E_1.$$

Algorithm 2 may be generalized as follow:

1. Compute $\hat{x}^{k+1}$ solution for some $\delta_k > \delta$ of

   $$(SP4) \qquad \min f(x)\ ;\ Ax = b,\ (x - x^k)^T M_k (x - x^k) \le \delta_k^2,$$

   where M_k is an appropriate positive definite matrix, δ_k being such that $B\hat{x}^{k+1} < d$ and $\ell < \hat{x}^{k+1} < u$.

2. Line search : denote $d_k := \hat{x}^{k+1} - x^k$.
   Compute $\rho_k = \text{argmin}\{f(x^k + \rho d^k),\ \rho \in [1, \gamma_k]\}$ where γ_k is such that $B(x^k + \gamma_k d^k) < d$ and $\ell < x^k + \gamma_k d^k < u$.

We may choose different matrices M_k such that they define an interior trust region : i.e.

$$\{x \in \mathbb{R}^n\ ;\ (x - x^k)^T M_k (x - x^k) < 1\} \subset \{x \in \mathbb{R}^n\ ;\ Bx < d,\ \ell < x < u\} \tag{4.20}$$

In our tests, we have taken $M_k = D_k^{-2} + B^T Z_k^{-2} B$ where

$$\begin{aligned} D_k &:= \text{diag}\{\min(x_i^k - \ell_i,\ u_i - x_i^k)\} \\ Z_k &:= \text{diag}\{(d - Bx^k)_i\}. \end{aligned}$$

It is easy to show that matrix M_k is positive definite and verifies (4.20). Now, the linear system to have to solve is similar to (4.19):

$$\begin{bmatrix} Q + \nu_k M_k & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} x_{\nu_k} - x^k \\ \lambda^k \end{pmatrix} = - \begin{pmatrix} \nabla f(x^k) \\ 0 \end{pmatrix} \tag{4.21}$$

Note that when the upper bound components are infinite, one can easily show that (SP4) is mathematicaly equivalent to sub-problem (SP3).

The advantage of working with the scaling matrix D_k defined above when variables are bounded by ℓ and u is that it ensures a larger ellipsoid than the one obtained after having used slack variables to transform the bound constraints to equality constraints. Indeed, using the scaling matrix D_k, the ellipsoid given by sub-problem (SP4) is

$$E_2 = \{x, (x - x^k)^t D_k^{-2} (x - x^k) + (x - x^k)^T B^T Z_k^{-2} B (x - x^k) \le \delta^k\}.$$

Therefore we see that E_2 strictly contains E_1.

Before going to numerical aspect of the algorithm, we briefly shall prove the convergence properties of the basic algorithm where the affine scaling is done using the matrix D_k. With no loss of generality, we focus on the particular problems

$$(P'') \quad \min f(x) \; ; \; Ax = b, \; \ell \leq x \leq u.$$

At each iteration of the basic algorithm we compute x^{k+1} solution of

$$\min f(x) \; ; \; Ax = b, \; (x - x^k)^T D_k^{-2} (x - x^k) \leq \delta^2.$$

As in section 2, we check that the algorithm is convergent :

Theorem 4.1 **i)** *Any accumulation point $\bar{x}$ of the sequence (x_k) is an optimal global solution of the reduced problem:*

$$\begin{cases} \min f(x) \\ Ax = b \\ x_I = u_I \\ x_J = \ell_J \end{cases}$$

where

$$\begin{array}{rcl} I & = & \{i \in \{2, \ldots, n\} / \bar{x}_i = u_i\} \\ J & = & \{i \in \{1, \ldots, n\} / \bar{x}_i = \ell_i\} \end{array}$$

ii) *If for any two subsets I and J of $\{1, \ldots, n\}$ such that $I \cap J = \emptyset$, we have that the problem*

$$\begin{cases} \min f(x) \\ Ax = b \\ x_I = u_I \\ x_J = \ell_J \end{cases}$$

has at most one solution, then the sequence (x^k) converges.

iii) *If the sequence (x^k) converges to $\bar{x}$ and if $\bar{x}$ is qualified then $\bar{x}$ satisfies the first-order optimality system of (P") i.e.,*

$$\begin{cases} [\nabla f(\bar{x}) + A^T \bar{\lambda}]_i = 0 & \text{if} \;\; \ell_i < \bar{x}_i < u_i \\ [\nabla f(\bar{x}) + A^T \bar{\lambda}]_i \geq 0 & \text{if} \;\; \bar{x}_i = \ell_i \\ [\nabla f(\bar{x}) + A^T \bar{\lambda}]_i \leq 0 & \text{if} \;\; \bar{x}_i = u_i \\ A\bar{x} = b \\ \ell \leq \bar{x} \leq u & . \end{cases}$$

iv) *If f is convex then any non degenerate limit point of (x^k) is an optimal solution of (P'').*

Proof: i) Noting that the multiplier ν^k goes to zero so that $\nabla f(x^{k+1})+A^T\lambda^{k+1} = \nu^k D_k^{-1}(x^k - x^{k+1})$, and from the definition of D_k, we deduce that $(\nabla f(x^{k+1}) + A^T\lambda^{k+1})_i \to 0$ where $i \notin I \cup J$. As in the proof of (ii) of theorem 2.1, it is easy to deduce i).

ii) Since $|x_i^{k+1} - x_i^k| \leq \delta \min(x_i^k - \ell_i, u_i - x_i^k)$ $\forall i = 1, \ldots, n$, we deduce with (i) that $\|x^{k+1} - x^k\| \to 0$; consequently the set of limit points of (x^k) is connected. By hypothesis this set is finite. It follows that $\{x^k\})$ converges.

iii) The fact that $\bar{x}$ is not degenerate implies that $\{\lambda^{k+1}\}$ converges to some $\bar{\lambda}$. If $\ell_i < \bar{x}_i < u_i$, we have that $(\nabla f(x^{k+1}) + A^T\lambda^{k+1})_i \to 0$. If $\bar{x}_i = \ell_i$, assume by contradiction that $(\nabla f(\bar{x}) + A^T\bar{\lambda})_i < 0$, hence $\nu^k(D_k^{-1}(x^k - x^{k+1}))_i < 0$ for all k large enough. Thus $\ell_i < x_i^k \leq x_i^{k+1}$ for all k large enough, which contradicts the fact that $\bar{x}_i = \ell_i$. As for the previous case, it is easy to show that if $\bar{x}_i = u_i$ then $[\nabla f(\bar{x}) + A^T\bar{\lambda}]_i \leq 0$.

iv) The proof is an immediate extension of the proof of (iv) of Theorem 2.1. □

The main aim of our numerical tests is to study the behaviour of Algorithm 2 in practice. We are not interested in the way to compute the initial point x_0. We also use the optimal value of the tested problem in the stopping test.

Specifically we compute the optimal value γ^* by an active set method (Casas and Pola [9]) and we stop algorithm 2 at iteration k when

$$|f(x^k) - \gamma^*| \leq 10^{-5}\frac{(f(x_0) - \gamma^*)}{(f(x_0) - \gamma^* + 1)}$$

We took the constant δ equal to 0.99 and ε the precision to stop the dichotomic procedure equal to 10^{-16} . The linesearch was exact and done in the direction d^k from $\hat{x}^{k+1}$ to 99 % of the way to the boundary of the feasible region.

All our tests were generated such that the point $e = (1, \ldots, 1)^t$ was the initial interior point to start the algorithm. Therefore, the right hand sides b and d in (P') were built such that :$Ae = b$ and $(Be)_j + j = d_j$. The components of the lower and upper bounds are made such that : $u(i) = i + 1$ and $\ell(i) = -(n + 2 - i)$. The matrices Q are generated by computing $Q = H^T H$ where H is a random (n, n) matrix.

At each test we have used two random acces modes to generate entries of A, B, H and C : a uniform random acces in $[0, 1]$ and a normal random acces with mean equal to zero and variance equal to one.

We report here the worst results of this statistics. To compute ν_{sup} in the algorithm we need an under estimate γ_k of $f(x^{k+1})$. In our tests, we first computed the optimal solution of the problem without inequality constraints and stopped if it is feasible. Otherwise, we took its function value as γ_k.

Since it is natural that the number of iterations of the algorithm decreases when the number of equality constraint increases, our tests reported here use only one equality constraint. However, we always use inequality and bound constraints. The figures below summarize the principal numerical results obtained :

Result 1 Figure 1 shows that without linesearch, algorithm 2 always converges in a reasonable number of iterations related to the sizes of the tested problems. In Figure 2, we see that the use of the exact linesearch allows to divide the number of iterations roughly by two.

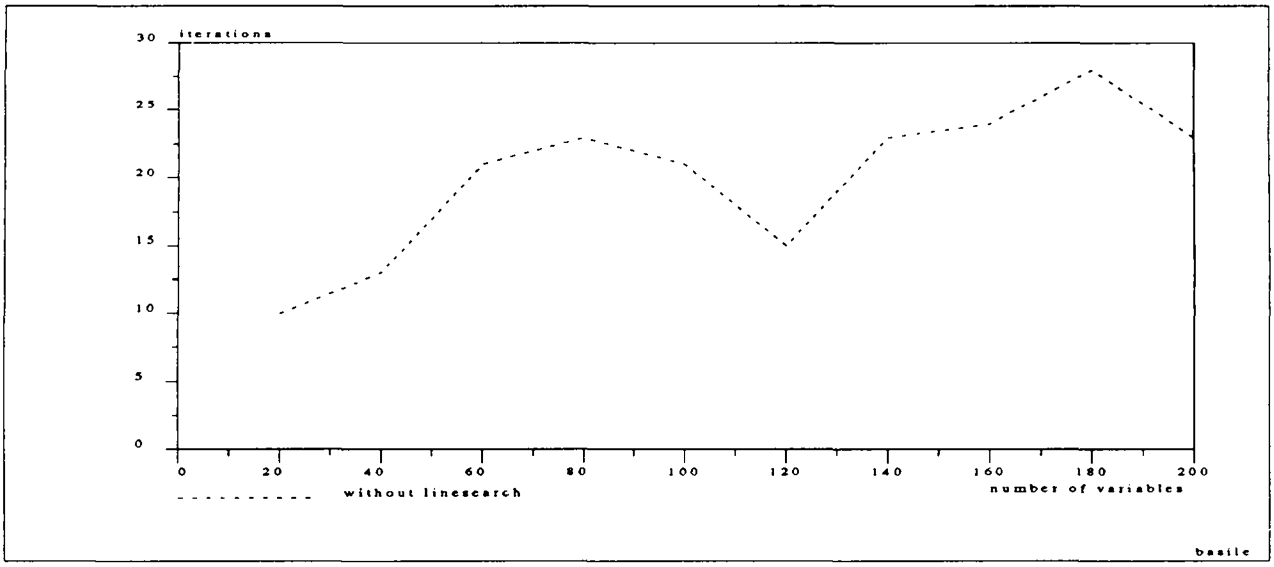

Figure 1: Number of inequality constraints fixed at 100. number of equality constraints equal to 1. Bound constraints for all variables.

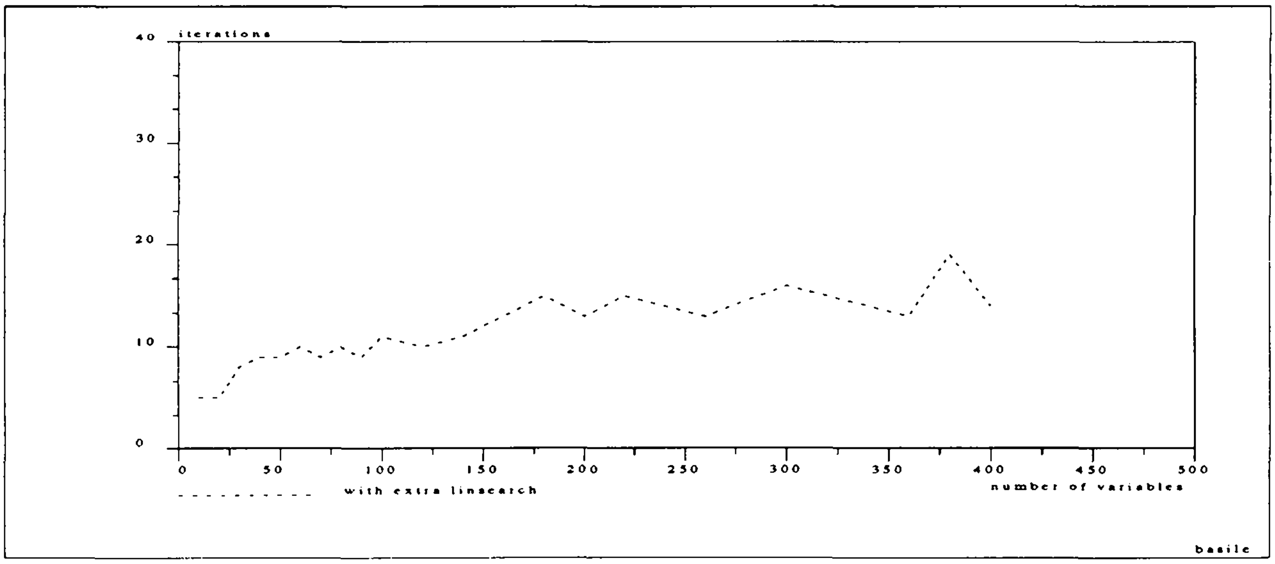

Figure 2: Same as in graphic 1 but with 200 inequality constraints.

Result 2 In Figure 3, the same function is minimized under a variable number of inequality constraints. We observe that by contrast to the classical active set methods, the speed of the algorithm is not very sensitive to the variation of the number of inequality constraints. This important behavior, is certainly one of the consequences of the fact that the algorithm operates inside of the feasible polyhedron.

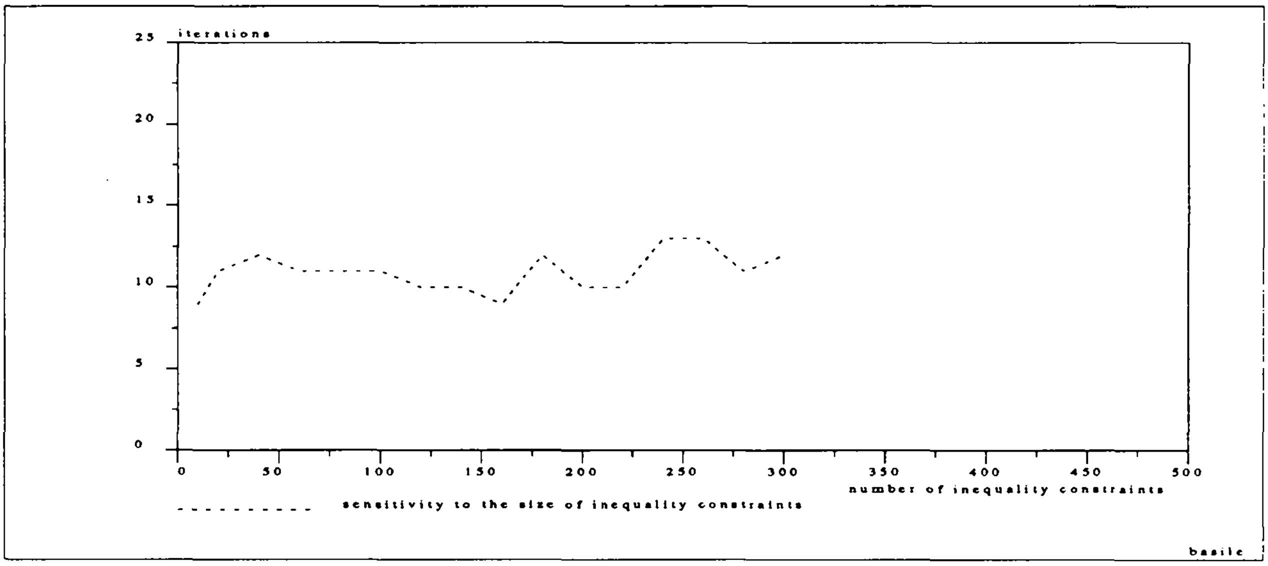

Figure 3: The number of variables is 100. One equality constraint. Bound constraints for all variables.

Result 3 In Figure 4 we show the importance of using ν^{k-1} as an upper estimate for the multiplier ν^k Indeed, the dotted line in the Figure 4, represents the case when we take the value $2\dfrac{f(x^k) - \gamma_k}{\delta^2}$ as the initial (upper) estimate on ν_k, and for the continuous line we compute ν^k in $[0, \nu^{k-1}]$. With this last version, we see that, before the convergence, each iteration usually needs to solve two linear systems and five when close to the convergence. This is very promising and we believe that with the best choice of the upper bound of ν^k and using some preconditioner for the matrix $Q + \nu_k M_k$, when close to the convergence, one can reduce the number of the linear systems to be solved.

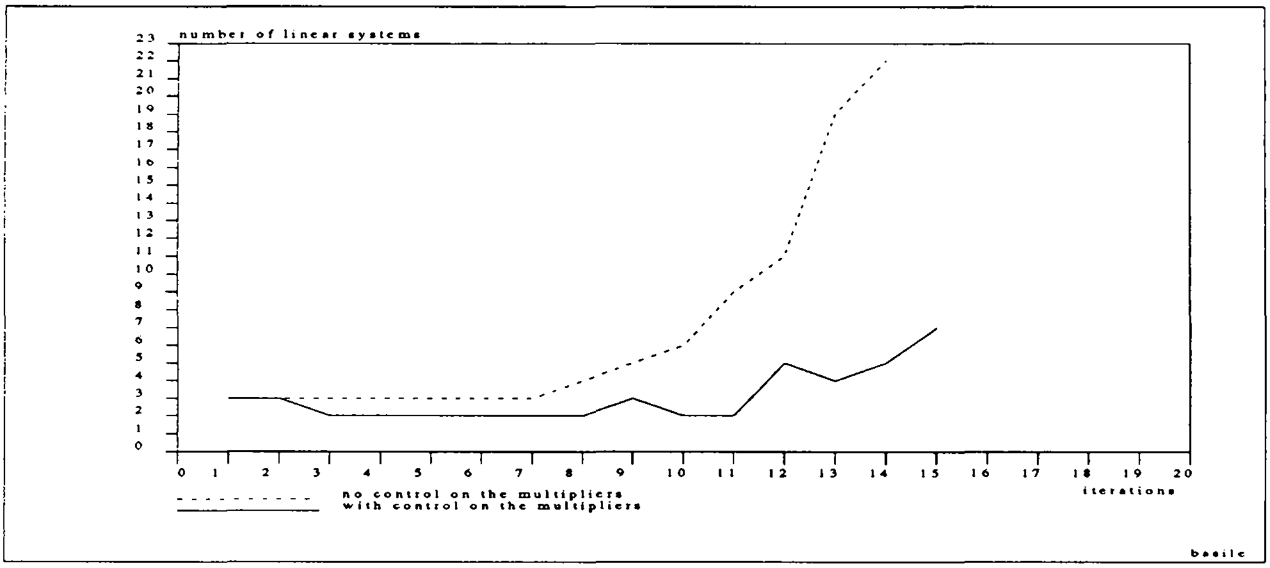

Figure 4: The number of variables is 100. One equality constraint. Bound constraints for all variables. 100 inequality constraints.

References

[1] I. Adler, M.G.C. Resende and G. Veiga : An implementation of Karmarkar's algorithm for linear programming. Math. Programming 44(1989), 297-335.

[2] I. Adler, R.D.C. Monteiro : Limiting behavior of the affine scaling continuous trajectories for linear programming problems. Math. Programming 50(1991), 29-51.

[3] E.R. Barnes : A variation on Karmarkar's algorithm for solving linear programming. Math Programming 36(1986), 174-182.

[4] F. Delebecque, C. Klimann, S. Steer : Basile. Un système de CAO pour l'automatique. Version 3.0. Guide d'utilisation. Mars 1989, INRIA.

[5] D.A. Bayer, J.C. Lagarias : The nonlinear geometry of linear programming : I Affine and projective trajectories. Trans. Amer. Math. Soc. 314(2), (1989), 499-526.

[6] M. Bouhtou : Méthodes de points intérieurs pour l'optimisation des problèmes de grande taille. Doctoral Thesis, Université Paris Dauphine.

[7] J.P. Bulteau, J.P. Vial : A restricted trust region algorithms for unconstrained optimization. Optim. Theory and Appl., Vol. 47, n° 4 (decembre 1985),413-435

[8] J. Burke, M. Ferris : Characterization of solution sets of convex programs. Operations Research Letters, 10 (1991), pp. 57-60.

[9] E. Casas , C. Pola, Dealing with degenerated cases in quadratic programs, RAIRO - Operations research, to appear.

[10] I.I. Dikin : Iterative solutions of problems of linear and quadratic programming, Soviet Math. Doklady 8(1967), 674-675.

[11] M. Gay : Computing optimal locally constrained steps. Siam J. Sci. Stat. Comput., Vol. 2, n° 2, (1982), pp. 186-197.

[12] C.C. Gonzaga , L.A. Carlos : A primal affine scaling algorithm for linearly constrained convex programs. Report, Federal University of Rio de Janeiro.

[13] C.C. Gonzaga : An interior trust region method for linearly constrained optimization, int. report, Federal University of Rio de Janeiro, Cx. P. 68511, 97945 Rio de Janeiro, RJ Brazil.

[14] C.C. Gonzaga : Convergence of the large step primal affine scaling algorithm for primal non-degenerate linear programs. Tech. Report, Federal University of Rio de Janeiro, Brazil.

[15] M.D. Hebden : An algorithm for minimization using exact second derivatives. Atomic Energy Research. Establishment report TP 515, Harwell, England 1973.

[16] N. Karmarkar : A new polynomial time algorithm for linear programming. Combinatorica 4(1984), 373-395.

[17] M. Kojima, S. Mizuno, A. Yoshise : An $0(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems. Math. Programming 50(1991), 331-342.

[18] O.L. Mangasarian : A simple characterization of solution sets of convex programs. Oper. Res. Lett., 7 (1988), 21-26.

[19] N. Megiddo and M. Shub : Boundary behavior of interior point algorithms for linear programming. Math. Oper. Res. 14, (1989), 97-146.

[20] R.D.C. Monteiro, I. Adler an M.G.C. Resende : A polynomial time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extensions. Math. Oper. Res. 15(2), 1990.

[21] J.J. Moré : Recent developments in algorithms and software for trust region method. Born 1982, Springer Verlag.

[22] C.H. Reinsch : Smoothing by spline functions, II, Numer. Math. 16, 451-454.

[23] D.C. Sorensen : Newton's method with a model trust region modification. SIAM J. Numer. Anal. 19, n° 2, 1982, pp. 409-426.

[24] J. Sun : A convergence proof of an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions. Technical Report, Department of Industrial Engineering and Management Sciences. Northwerstern University, Evanston, IL 60208, USA, 1990.

[25] P. Tseng and Z.Q. Luo : On the convergence of the affine scaling algorithm. Technical Report, CICS-P-169, Center for Intelligent Control Systems, Massachusetts Institute of Technology (Cambridge, MA, 1989).

[26] T. Tsuchiya : Global convergence of the affine scaling methods for degenerate linear programming problem. Math. Programming 52, (1991), pp. 377-404.

[27] T. Tsuchiya : Global convergence of the affine scaling algorithm for the primal degenerate strictly convex quadratic programming problems. The Institute of Statistical Mathematics, Tokyo, Japan, Technical report n°417.

[28] T. Tsuchiya and R.D.C. Monteiro : Superlinear convergence of the affine scaling algorithme. Department of Systems and Industrial Engineering. University of Arizona, Tucson, Technical Report.

[29] R.J. Vanderbei, S.M. Meketon, B.A. Freedman : A modification of Karmarkar's linear programming algorithm. Algorithmica (1986) 1, 395-407.

[30] Y. Ye, E. Tse : An extension of Karmarkar's projective algorithm for convex quadratic programming. Math. Programming 44 (1989), 157-179.

[31] Y. Ye : On affine scaling algorithms for nonconvex quadratic programming. Math. Prog. 56(3), 1992, pp. 285-301.

Unité de Recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)
Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399

RR-1947